EPISCIENCES - an overlay publication platform

Christine Berthaud, Laurent Capelli, Jens Gustedt, Claude Kirchner, Kevin Loiseau, Agnès Magron, Maud Medves, Alain Monteil, Gaëlle Riverieux, Laurent Romary

Let's Put Data to Use: Digital Scholarship for the Next Generation
P. Polydoratou and M. Dobreva (Eds.)

doi:10.3233/978-1-61499-409-1-78

EPISCIENCES – an overlay publication platform

Christine BERTHAUD[a], Laurent CAPELLI[a], Jens GUSTEDT[b,c], Claude KIRCHNER[b], Kevin LOISEAU[a], Agnès MAGRON[a], Maud MEDVES[b,d] , Alain MONTEIL[b], Gaëlle RIVERIEUX[b], Laurent ROMARY[b,d,1]

[a] *Centre pour la Communication Scientifique Directe (CCSD), CNRS : UPS2275*
[b] *Inria, France*
[c] *Université de Strasbourg, ICube, Illkirch, France*
[d] *Humboldt-Universität zu Berlin, Institut für Deutsche Sprache und Linguistik (IDSL)*

Abstract. This paper delineates the main characteristics of the Episciences platform, an environment for overlay peer-reviewing that complements existing publication repositories, designed by the Centre pour la Communication Scientifique directe (CCSD[2]) service unit. We describe the main characteristics of the platform and present the first experiment of launching two journals in the computer science domain onto it. Finally, we address a series of open questions related to the actual changes in editorial models (open submission, open peer-review, augmented publication) that such a platform is likely to raise, as well as some hints as to the underlying business model.

Keywords. Overlay journal – Editorial platform – Scholarly communication-Repositories – Open Access

1. Exploring new scholarly publication models

The recent debates on Open Access have mainly focused on opposing models, the so-called green model, where scientists deposit their (possibly published) research papers in open repositories and the gold model where publishers, usually following the payment of an author fee, freely release the publication online. This debate often misses two points. First, that what is at stake is to have a reliable and sustainable communication system for science where scientists themselves have the say and are provided with all means to quickly disseminate their results while receiving the appropriate feedback (usually embodied by peer-reviewing) from their communities. Second, that all data generated around the evaluation, the reviews and the associated discussions (forums, etc.) shall be monitored by the scientific community.

Still, we know that alternative models to the traditional publisher-owned journals are possible, and experiences carried out in the human sciences with the OpenEdition endeavour for instance have shown that research communities may react favourably

[1] Corresponding author
[2] CCSD is a joint service unit between the CNRS, Inria and the University of Lyon

when a real alternative is being offered. Such initiatives provide a systemic concept of publishing (from scholarly blogs to journal publications) comprising both new editorial frameworks and business models.

In this context, we present a new initiative to provide an overlay journal environment, i.e. a journal that is built as an additional peer-reviewing layer on top of a publication repository (see Smith, 1999). This environment offers a technical and editorial platform for existing or new journals operated within a multi-institutional and publicly controlled infrastructure based upon a large-scale publication archive. By sharing the technical settings with a publication repository and focusing on the core missions of a scientific journal we expect to both reduce costs dramatically and open possibilities of experimenting new certification mechanisms.

To quote 0: "The underlying vision is that of a research infrastructure where no fee is applied to its users (whether author or reader) and which offers a set of basic services facilitating an efficient dissemination and review of scholarly papers. Like traditional journals, scientific quality is ensured by the recognition of the editorial committee that carries out the peer-reviewing process." Part of the uniqueness of the Episciences endeavour is the strong commitment of national institutions in ensuring both the quality of the service and its anchoring within a sustainable infrastructure.

In the remaining sections of this paper we will first show how an overlay journal is homothetic to the traditional journal publication principles. We will then describe the role of the publication archive in providing a set of core services for the deployment of a peer-reviewing environment and see what additional functionalities have been designed for the Episciences platform. We will identify which core mechanisms are required to provide a reliable certification service and which may be more peripheral. Finally we will present the first experiment carried out while launching two journals, namely DMTCS (Discrete Mathematics & Theoretical Computer Science) and JDMDH (Journal of Data Mining and Digital Humanities)[3], onto the platform and discuss various topics related to the potentialities offered by overlay journals.

2. Overlay journals seen as a specific case of scholarly journals

2.1. The main functions of a scholarly publishing platform

In his 2009 and 2010 papers, M. Mabe outlines the role of scholarly publishing along the following dimensions:

- *Registration*: the process of submitting a paper, which establishes the author's precedence and ownership of an idea
- *Certification*: where quality control is ensured through peer-review, and consequently scholarly reward is provided to the author
- *Dissemination*: the communication of the findings to its intended audience

[3] Jointly launched by the two French research organizations INRA (Agronomics) and Inria (computer science)

- *Archival record*: preserving a fixed version of an article for future reference and citation.

Whereas this description nicely and conservatively describes the current publisher based setting of scholarly publishing, it may be subject to discussion when considering which new models should be experimented or further deployed.

A first element of discussion is whether all four functions should be situated within the same platform to be fulfilling the researchers' expectations. For instance, managing trustful affiliations is typically part of the competence of a research institution rather than that of a publisher. In the same way, archiving and managing a reference corpus of scholarly papers may be part of the core missions of a community, as exemplified by the initiatives carried out by scholarly associations such as the Association for Computational Linguistics (ACL) or the Association for Computing Machinery (ACM). Finally, it is easy to imagine that certification and dissemination can be completely disconnected from one another in a mediated world where social networks are more and more used to convey daily scientific news.

More importantly, we can see how this frozen scenario may be counter-productive to the very essence of scholarly publishing, namely to ensure the appropriate convey of knowledge between scholars, but also to the wider public. First, it subordinates the dissemination of scholarly papers to the peer-reviewing process, whereas we know how much the two can live independently from one another (see Gentil-Beccot et alii, 2009), but also how much danger there is when a selective review process prevents the dissemination of useful results[4]. This situation leads scholars to submit their papers iteratively to multiple settings and reviewers to get drowned under a deluge of useless refereeing work.

The whole idea of the Episciences initiative is to decompose the process to ensure maximal efficiency at the service of scholarly communication. In particular, we now see how publication repositories can play a core role for an open publication process.

2.2. Publication repositories as an infrastructure for scholarly publishing

Open archives are now widely available and can be used by any researcher to store, index and make any of their research documents freely available, whether or not these have been published in peer-reviewed channels (journals or conferences). Even more, these documents can range from research papers to experiments, data, computer programs or videos. Such archives as the e-print archive arXiv or Hyper Articles en Ligne (HAL) are widely accessible and provide a free and sustainable service. In the case of the HAL platform for instance, papers are associated with precise affiliation information for each author, and are supported by long-term archiving facilities. Additional services like the creation of personal or institutional web pages are also offered.

Seen from the point of view of scholarly publishing we can see how most existing publication archives provide an adequate environment for supporting several of the core functions related to traditional journals (see Romary & Armbruster, 2010):

[4] See for instance:
http://www.nature.com/news/half-of-us-clinical-trials-go-unpublished-1.14286

- They provide a reliable registration environment whereby both attribution (authors and their affiliations) as well as time-stamping[5] are attached to the registered documents;
- Dissemination is naturally ensured not only through the built-in open access nature of the archive but also because large scale publication repositories such as HAL[6] or arXiv are highly visible within search engines and their content are followed (alert mechanisms) by the research communities;
- Finally, archival record is also a natural component of publication repositories, with an additional advantage here, namely that papers from a given author or institution can be gathered within a coherent setting rather than being spread across various publishers' portals, whose long-term existence or accessibility is far from being ensured.

Beyond these standard functionalities, institutional publication archives often come with various additional features that make them even more powerful than usual publishers' environments. First, being hosted by sustainable institutions, they offer some guaranties that the technical environment and thus the corresponding content will be made available for a long period of time. This is even more the case for central repositories such as HAL, where a consortium of institutions, or even a national policy[7], is backing up the service. Research libraries also often curate the content, thus ensuring coherent metadata descriptions associated with authority lists of institutions or funded projects.

From a technical point of view, it is also important to apprehend how much versioning is an essential feature from the point of view of the academic process since it allows researchers to trace the processes when writing a document and, possibly, integrating the comments received from their colleagues, anonymously or not.

As a whole, we see that only a core set of mechanisms have to be implemented to fulfil the role of a scholarly journal environment, namely a) the management of the review process and b) the provision of more or less fine-grained copy-editing support. The following sections will describe how the Episciences project fulfils these.

3. Main functions of the Episciences publication platform

The Episciences platform is conceived in the spirit of traditional peer-reviewed journals, with additional facilities resulting from it leaning against a publication repository. The editorial team and the reviewing and publication workflow are standard, with the difference that the paper is managed by the author and not by the editors in charge, the labelling of the paper as accepted being of course fully handled under the control of the editorial board. This impacts on copy-editing because the

[5] This is for instance essential if these documents are to be used to assess the anteriority of a discovery as is the case for prior art search in patent organizations.

[6] HAL was ranked 5th in the Webometrics portal ranking as of Jan. 2014 (see http://repositories.webometrics.info/en/top_portals)

[7] see http://www.enseignementsup-recherche.gouv.fr/cid71277/partenariat-en-faveur-des-archives-ouvertes-plateforme-mutualisee-hal.html (in French)

author is responsible for the layout (unless he gives over some rights) and versioning with all versions of the paper (at least the submitted and accepted ones) being available on the repository.

3.1. Editorial services

In order to support the editors-in-chief and editorial boards in their day-to-day business, a support in terms of editorial management is provided. This comprises:

- Management of the peer-review process, comprising the channelling of community based feedback and the plagiarism detection;
- Handling the management of the journal volumes and issues;
- Contribution to some basic quality checking tasks (bibliography, metadata, cross-references, automatic detection of the state of the art);
- Communication and community management: advertising journals and papers through various channels and social networks (twitter, blog, academic social website), moderation of online discussions (made possible by the commenting functions and display of tweets related to an article)[8];
- General visibility: interaction with major indexing services and databases (Digital Bibliography and Library Project, Thomson Reuters, Scopus...), as well as adequate mirroring on relevant thematic repositories (ArXiv, PubMed Central, Research Papers in Economics, etc.).

3.2. Technical services

Through the hosting on the French national repository infrastructure HAL, all journals benefit from a high quality technical environment comprising 24/7 services, long term archiving of all versions and proper authentication and authorisation infrastructure. Other platforms such as arXiv offer similar facilities.

The platform offers web design tools so that each journal can customise its own website while their generic graphical identity retains features of the Episciences design.

Long term archiving of the reviewing information is also assured: the ratings as well as the exchanges between authors and reviewers are securely stored on the platform and are accessible to the editorial team at any time. According to the journal policy, reviews may be published as well as the reviewer's names (see discussion in section 6).

3.3. Intellectual property management

The Episciences model impacts at several levels on intellectual property issues. First, the Episciences platform leaves all rights to the journals concerning the ownership of the title. The basic idea here is that the platform will not be the publisher. In cases

[8] Such services have already been experimented for HAL: cf. http://fronthal.imag.fr/noModule

where there may be difficulties to manage such an ownership[9], the consortium of institutions in charge of Episciences will upon request temporarily host the ownership of the title.

From an author's point of view, a simple non-exclusive licence will be requested. As a matter of fact, given that the papers are available through a publication archive, they actually bear the associated open licence (in the case of HAL-Inria for instance a strong recommendation is made to have papers issued under the Creative Commons Attribution (CC-BY) licence).

3.4. Copy editing

Copy-editing is left to the editorial board of each journal, which will also decide of the submission format and style. Typically, submission in TeX or LaTeX may ensure that the formatting instructions will be slightly better met in most cases without any need for further copy-editing related to the actual formatting of papers. Still, we are aware that copy-editing is a question. The quality that is provided by author sources is very much varying, and there is not only a quality control job involved, but many authors definitively need help and guidance, and for some much of the work may have to be provided. Part of the developments we will have to consider (see section 6 below on the budget break-out) is to be able to support journals with such needs.

4. Managing the Episciences journal portfolio

The journals hosted on the Episciences platform are organised as thematic portfolios. The objective is to ensure quality and coherence on a discipline based rationale. In order to achieve this, each scientific domain that will have journals on Episciences will form a pool coordinated by a so-called meta-committee, a group of internationally recognised experts whose duty will be to select new incoming journals, check out their overall operation and quality, but also be the contact to attract new journals within their respective communities. Part of the duties of a meta-committee will also be to control the thematic coherence of the various journals, so that clear guidance can be given to authors as to where their papers should be optimally submitted.

Two such meta-committees are currently being set-up in Mathematics and Computer Science, which correspond to the communities that have started to show interest for Episciences.

5. Two initial experiments

We started the platform with two journals from different sub-domains in computer science. One of the journals, JDMDH[10], is a new creation, corresponding to an emerging domain with a scientific committee that has collectively decided to go for an open journal and to join efforts with Inria on the new platform. The other one,

[9] when not properly hosted by an academic institution or a scientific society

[10] Note that no official launch has been made yet at the time of submission of the paper and that the site is still in test phase

DMTCS, is an established open journal for which we designed a transition scheme to Episciences.

JDMDH covers all aspects of data mining methods for the humanities. The first launch issue is in preparation with all submitted papers already deposited in the Episciences framework (namely deposited on HAL and arXiv prior to submission to the journal). There is already a strong support within the editorial committee for the post-publishing peer-review process (see also discussion in section 6).

DMTCS is a well-established scientific journal. Placed at the cross-section between computer sciences and mathematics, it covers both, but emphasizes on work that profits for or from both. In the late nineties, DMTCS was one of the first open access journals that came to life, in a then rapidly growing context of the still new and chilling Internet. At first managed by a commercial editing house, the DMTCS title was quickly transferred to the scientific editors. DMTCS is structured in volumes and issues, though they are only formal remainders of ancient publishing traditions. De facto the journal is published continuously.

The online system[11] evolved from a collection of simple web pages and an editorial process managed through mail, over a home-brew server software, to the Open Journal System (OJS). Without dedicated specialised staff, the journal is clearly vulnerable and lacks reactivity and quality of service.

One of the main challenges when migrating DMTCS from OJS to Episciences was to manage legacy papers. First, it was necessary to keep two platforms alive in parallel for a while, namely until the peer-review process of the articles submitted in OJS is over (while new articles are submitted in Episciences). Second, it proved challenging to import all legacy papers into HAL with the expected level of metadata precision.

6. Issues raised by an overlay journal platform

The Episciences model is not a simple replacement of the traditional scholarly publishing environment. Its integration within the services of publication repositories in particular makes it bear specific characteristics, which we would like to analyse in this section, being aware that many consequences of the model are likely to appear when processing a larger portfolio of journals.

6.1. A low-cost platform

The economic study[12] of the EU-funded Publishing and the Ecology of European Research (PEER) project evaluated (p.48) the cost in a repository to range between 2 and 50 € per reference and between 2,5 and 53,2 € per full text[13]. It also showed that a baseline for managing the peer-review process alone lies around 200 € per article for most commercial journals. Such costs usually correspond to the manpower related to editorial secretariat and is planned to be one of the possible duties of future librarians within research institutions, as anticipated in (Guédon, 2001).

---

[11] With the support of Inria and Loria laboratory

[12] http://www.peerproject.eu/fileadmin/media/reports/PEER_Economics_Report.pdf

[13] Note that for HAL the average cost per paper has been evaluated to 14.73€

For such cost we need to be open as to the possible business models that may allow our initiative to break even in the long run. We basically see three main possible components for a balanced funding scheme:

- Following the model adopted for HAL, we have started to pool some core resources within a consortium of partners. The stability of such national institutions will ensure sustainability for the platform;
- We also need to unite forces with initiatives such as OpenEdition which sell additional services (cataloguing, smart formats (ePub)) to university libraries, whose benefits directly finance the journals themselves;
- We should not reject author processing charges when there is a request for additional copy-editing services, such as suggested by the Copernicus publisher for its open access journals.

6.2. Leaving away the post peer-review publishing paradigm

One important consequence of the overlay journal model is that papers are made public right at the time of their deposit on the publication repository, which means that the peer-review process actually takes place *after* the actual publication[14]. There are several consequences that derive from this principle:

- Having the paper online before peer-review obviously prevents author anonymity. Whereas this is not necessarily part of the cultural background of some scholarly communities, there are strong arguments to see this as a benefit for the scholarly process (see 0 and next section on open peer-review)
- Whatever the time and the duration of the review process, the paper benefits from a high visibility right from the onset. This may allow colleagues to comment at an early stage and even for the document to be cited if already relevant as background for another research. This aspect has become normal practice for many communities like in physics or astronomy with arXiv as a pre-print server;
- The paper remains available whatever the success of the peer-review, which guaranties the continuous availability of the corresponding results independently of the outcomes and possibly incidents of the certification process. This is important to circumvent the dramatic loss on non-published information that science currently faces (see Jones et alii, 2013);
- The experience gained from other open reviewing environment (see Pöschl, 2004) has shown that open manuscripts reduce the number of poorly written submissions, thus leading to a more efficient peer-review process;
- The paper may evolve further if new elements validating or invalidating the paper are discovered. An overlay publication system thus facilitates the management of versions (or errata in the mathematical domain).

[14] See the position blog entry by J. Velterop: http://theparachute.blogspot.co.uk/2013/11/essence-of-academic-publishing.html

The issue at stake is how much such a model will be accepted by a variety of scholarly communities or if we may have to allow "invisible" papers in publication archives to cover more publication scenarios.

6.3. Towards new peer-review models

Once the psychological barrier of post peer-review publication has been overcome, a platform such as Episciences is the ideal place to convince scientific communities that peer-review can take other forms than those known in traditional journal settings. There are indeed two complementary directions that we would now like to pursue:

- Open peer-review, whereby reviews become openly accessible with, possibly, the identification of the reviewers. By doing so, we encourage reviews to become publication objects of their own and be part of a publication bundle together with the paper itself;
- Community feedback: by linking papers to scholarly blog entries or pushing submissions to external reviewing platforms (e.g. PeerEvaluation) to offer further commenting environments.

6.4. Towards new documentary services

Linking a journal platform to a national publication repository opens up a wide range of potential services that would not be affordable for such a dedicated peer-review platform. In the context of our current developments on the HAL platform, such services include automatic PDF to metadata recogniser[15] (title, author, affiliation, keywords and abstract information) to simplify the submission process for an author, or the automatic detection of bibliographical references for linking the paper to other relevant publications.

An important disruptive step will be to systematically create a reference XML version of all papers[16], which in turn can be used to produce different publication formats (HTML, ePub, PDF with a specific layout, etc.).

6.5. Episciences for putting together data journals

Finally, we can see that the Episciences workflow is designed independently of the nature of the initial document. It may indeed not be a textual object but a compound of notes, programs (possibly active) and data that could benefit from the same kind of certification process. The way towards data journals, which only a handful of communities have tackled so far, can be part of the realm of overlay certification processes, when anchored on data or program repositories[17].

---

[15] Based on https://github.com/kermitt2/grobid

[16] Compliant with the TEI guidelines (cf. http://www.tei-c.org)

[17] In the computer science domain, the IPOL journal (http://www.ipol.im) for instance deals with the assessment of executable computer programs.

7. Overview

We think that putting together such a platform for overlay journals, and making it widely available to research communities, will offer a whole wealth of features for scholars by providing fast and efficient dissemination of scholarly results. Beyond the maths and informatics communities that are now involved in this endeavour, we expect a wider range of domains to benefit from this service.

The experiment carried out with our two initial journals has allowed us to secure most of the features on the platform and validate that a quick, and cheap, deployment of an overlay journal is possible. We can now identify our roadmap for the future in two complementary directions: bring in more journals in the informatics and applied mathematics domain, where we have already felt a strong demand, and attract a wide range of interested institutions to join efforts in securing the long-term sustainability of the endeavour.

References

Gentil-Beccot A., S. Mele and T. C. Brooks (2009) "Citing and Reading Behaviours in High-Energy Physics. How a Community Stopped Worrying about Journals and Learned to Love Repositories." CoRR abs/0906.5418

Guédon J.-C. (2001), "In Oldenburg's Long Shadow: Librarians, Research Scientists, Publishers, and the Control of Scientific Publishing", *Association of Research Libraries*, ISBN 0-918006-81-3

Jones C. W., L. Handler, K. E. Crowell, L. G. Keil, M. A. Weaver, T. F. Platts-Mills (2013) "Non-publication of large randomized clinical trials: cross sectional analysis". *BMJ* 2013;347:f6104

Mabe M. A. (2009). "Scholarly Publishing". *European Review*, 17, pp 3-22. doi:10.1017/S1062798709000532

Mabe M. A. (2010) Scholarly Communication: A Long View, New Review of Academic Librarianship, 16:S1, 132-144, doi:10.1080/13614533.2010.512242

Pöschl, U. (2004) "Interactive journal concept for improved scientific publishing and quality assurance", *Learned Publishing*, 17(2), pp. 105-113.

Romary L., Armsbruster, C (2010). Beyond institutional repositories, in International Journal of Digital Library Systems **1**, 1, pp.44-61 http://hal.archives-ouvertes.fr/hal-00399881

Romary L.,, P. Guitton & C. Kirchner (2013) "A French perspective on Open Access and the Episciences Initiative", in T. Hey (Ed.) *Tony Hey on eScience* (tonyhey.net), June 2013, http://tonyhey.net/2013/06/03/a-global-view-of-open-access-part-1/

Smith, John W T "The deconstructed journal : a new model for academic publishing". *Learned Publishing*, 1999, vol. 12, n. 2.